



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2019

The Effect of Stimulus-Response Compatibility on the Association of Fluid Intelligence and Working Memory with Choice Reaction Times

Hülür, Gizem ; Keye-Ehing, Doris ; Oberauer, Klaus ; Wilhelm, Oliver

Abstract: It is a well-replicated finding that reaction time is correlated with performance in intelligence tests. According to the binding hypothesis of working memory capacity, the ability to establish bindings between elements and to integrate them into new structural representations is the source of the common variance between different cognitive tasks, including fluid intelligence and working memory. The goal of this study was to examine the effects of stimulus-response compatibility on the association between reaction time, fluid intelligence, and working memory. Based on the binding hypothesis, we expected that correlations between reaction time and fluid intelligence would be larger for arbitrary than for compatible stimulus-response mappings. We report data from two studies (Study 1: $n = 135$, mean age = 18 years; Study 2: $n = 153$, mean age = 17 years). We used choice reaction time tasks with compatible and arbitrary mappings as well as indicators of fluid intelligence and working memory (Study 1) and fluid and crystallized intelligence (Study 2). In both studies, we established a measurement model that included a factor reflecting general reaction time, and a nested factor reflecting the cost of establishing and maintaining arbitrary stimulus-response bindings. The results of Study 1 supported the hypothesis that the ability to uphold arbitrary bindings is correlated with working memory, but it was not correlated with fluid intelligence. In Study 2, the correlations between the binding factor and fluid and crystallized intelligence were again not significantly different from 0. We discuss possible reasons for these findings.

DOI: <https://doi.org/10.5334/joc.66>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-171519>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Hülür, Gizem; Keye-Ehing, Doris; Oberauer, Klaus; Wilhelm, Oliver (2019). The Effect of Stimulus-Response Compatibility on the Association of Fluid Intelligence and Working Memory with Choice Reaction Times. *Journal of Cognition*, 2(1):1-19.

DOI: <https://doi.org/10.5334/joc.66>

DATA REPORT

The Effect of Stimulus-Response Compatibility on the Association of Fluid Intelligence and Working Memory with Choice Reaction Times

Gizem Hülür¹, Doris Keye-Ehing², Klaus Oberauer¹ and Oliver Wilhelm³¹ University of Zurich, CH² University of Applied Labour Studies, DE³ Ulm University, DECorresponding Author: Gizem Hülür (gizem.hueluer@uzh.ch)

It is a well-replicated finding that reaction time is correlated with performance in intelligence tests. According to the binding hypothesis of working memory capacity, the ability to establish bindings between elements and to integrate them into new structural representations is the source of the common variance between different cognitive tasks, including fluid intelligence and working memory. The goal of this study was to examine the effects of stimulus-response compatibility on the association between reaction time, fluid intelligence, and working memory. Based on the binding hypothesis, we expected that correlations between reaction time and fluid intelligence would be larger for arbitrary than for compatible stimulus-response mappings. We report data from two studies (Study 1: $n = 135$, mean age = 18 years; Study 2: $n = 153$, mean age = 17 years). We used choice reaction time tasks with compatible and arbitrary mappings as well as indicators of fluid intelligence and working memory (Study 1) and fluid and crystallized intelligence (Study 2). In both studies, we established a measurement model that included a factor reflecting general reaction time, and a nested factor reflecting the cost of establishing and maintaining arbitrary stimulus-response bindings. The results of Study 1 supported the hypothesis that the ability to uphold arbitrary bindings is correlated with working memory, but it was not correlated with fluid intelligence. In Study 2, the correlations between the binding factor and fluid and crystallized intelligence were again not significantly different from 0. We discuss possible reasons for these findings.

Keywords: Response speed; Working memory; Reasoning

Intelligence measures correlate moderately negative with reaction time (RT) on relatively simple tasks (Danthiir, Roberts, Schulze, & Wilhelm, 2005; Jensen, 1993). One proposal for explaining this well-replicated finding arises from the binding hypothesis of working memory (WM) (Oberauer, 2005): Individuals with higher WM capacity are better in maintaining temporary bindings between representations. This enables them to build more complex structural representations in reasoning tasks. It also enables them to build more robust bindings between stimulus and response categories in speeded choice tasks. Accordingly, the ability to maintain robust temporary bindings is the common cause explaining the correlation between fluid intelligence (gf) and choice RT.

Wilhelm and Oberauer (2006) tested this idea by examining correlations between RT in choice tasks of varying stimulus response (SR) compatibility, WM and gf. They hypothesized that WM is involved in establishing and maintaining bindings between SR representations, and that such bindings are more important for non-compatible than for compatible SR mappings. Therefore, WM and gf should be correlated more strongly with RT in non-compatible than in compatible choice tasks.

Compatible SR mappings have been consistently linked to faster RT and higher accuracy than non-compatible mappings (e.g. Fitts & Deininger, 1954; Fitts & Seeger, 1953). According to the model of Kornblum, Hasbroucq, and Osman (1990), dimensional overlap is crucial for SR-compatibility. For instance, there is

dimensional overlap between stimuli and responses if both have spatial features. A task may be considered compatible or non-compatible depending on whether stimuli and responses have the same values on this shared dimension. For instance, if a stimulus on the left is mapped to a left response key, and a stimulus on the right to a right response key, then their mapping is compatible; the reverse mapping is non-compatible.

Non-compatible mappings can be incompatible or arbitrary. Incompatible mappings are obtained by reversing compatible mappings (e.g., left stimulus mapped to right response key and vice versa). In arbitrary mappings, there are no preexisting associations between SR sets (e.g., left stimulus mapped to upper response key, and right stimulus mapped to lower response key) or no dimensional overlap between stimuli and responses (e.g., a red light mapped to right key, and a green light mapped to left key).

In compatible tasks, SR bindings are established partly through preexisting associations in long-term memory. In non-compatible tasks, instructed SR mappings must rely exclusively on ad-hoc bindings in WM. Therefore, WM and gf should correlate higher with RT in non-compatible than in compatible tasks. To test this hypothesis, Wilhelm and Oberauer (2006) used four-choice RT tasks with compatible, incompatible, and arbitrary SR mappings with visual and auditory material respectively. Stimuli for the SR tasks varied on two dimensions: on one dimension that was relevant to the task and on another dimension that was irrelevant. In the visual task, the relevant dimension was location and the irrelevant dimension was color: Squares appeared at one of four possible locations horizontally arranged in a row and had one out of four colors (red, blue, yellow, or green). In the auditory task, the relevant dimension was the location word and the irrelevant dimension was the speaker who spoke the location words. Through headphones, participants listened to location words (“above”, “below”, “left”, or “right”) spoken by four different speakers. In the compatible condition, each stimulus was mapped to the corresponding response key. In the incompatible condition, the compatible mapping was reversed (e.g., upper response key assigned to the word “below”). In the arbitrary condition, the assignment of stimuli to response keys did not follow any obvious rule. The tasks are illustrated in **Figure 1**. Individual differences in average RT per condition were captured by two factors: A general RT factor reflecting individual differences in general speediness and a nested binding factor reflecting additional binding costs in arbitrary conditions. Correlations of the binding factor with WM ($\rho = -.89$) and gf ($\rho = -.55$) were higher than correlations of the general RT factor with WM ($\rho = -.53$) and gf ($\rho = -.42$). These findings are in line with the binding hypothesis.

Correlations between gf, WM, and choice RT with arbitrary and non-arbitrary SR mappings have also been examined by Meiran and colleagues in two different studies (Meiran, Peveg, Givon, Danieli, & Shahar, 2016; Meiran & Shahar, 2018). In the study by Meiran et al. (2016), non-arbitrary RT was significantly correlated with WM ($\rho = .22$) and arbitrary RT was not ($\rho = .12$). Both correlation coefficients were lower than those reported by Wilhelm and Oberauer (2006). In contrast, both arbitrary ($\rho = .72$) and non-arbitrary RT ($\rho = .59$) showed higher correlations with gf. Meiran and Shahar (2018) examined associations between gf, WM, and the tau parameter of the Ex-Gaussian model of RT distributions, reflecting the rate of exceptionally slow RTs in arbitrary and non-arbitrary tasks. gf correlated more strongly with the tau parameter of arbitrary ($r = -.64$) than of non-arbitrary ($r = -.45$) tasks. This finding is in line with the prediction of the binding hypothesis. WM correlated with neither tau parameter ($r = -.27$ for arbitrary and $r = -.11$ for non-arbitrary tasks). In contrast; Unsworth, Redick, Spillers, and Brewer (2012) found that individuals high vs. low in WM primarily differed in the RT of slowest responses in choice RT tasks, which they interpreted as lapses of individuals low in WM related to the active maintenance of goals.

Although the findings of Wilhelm and Oberauer (2006) support the binding hypothesis, they also allow an alternative explanation: There was dimensional overlap between SR sets for arbitrary mappings (see **Figure 1**): The corresponding compatible response key was among the response alternatives, but not the correct response. This may have resulted in response conflicts between the compatible and the correct response. Therefore, it cannot be ruled out that their finding was driven by higher response conflict and not binding requirements.

The present study

We aimed at expanding the findings of Wilhelm and Oberauer (2006) using different choice RT tasks: Response conflicts were avoided by generating SR sets without dimensional overlap in the arbitrary conditions. We examined the effect of SR-compatibility on correlations between RT, gf, and WM in two studies. We aimed at establishing a measurement model including a latent factor reflecting binding costs, that is, the additional time cost in arbitrary conditions compared to compatible conditions. According to the binding hypothesis, the binding factor should be correlated with gf and WM.

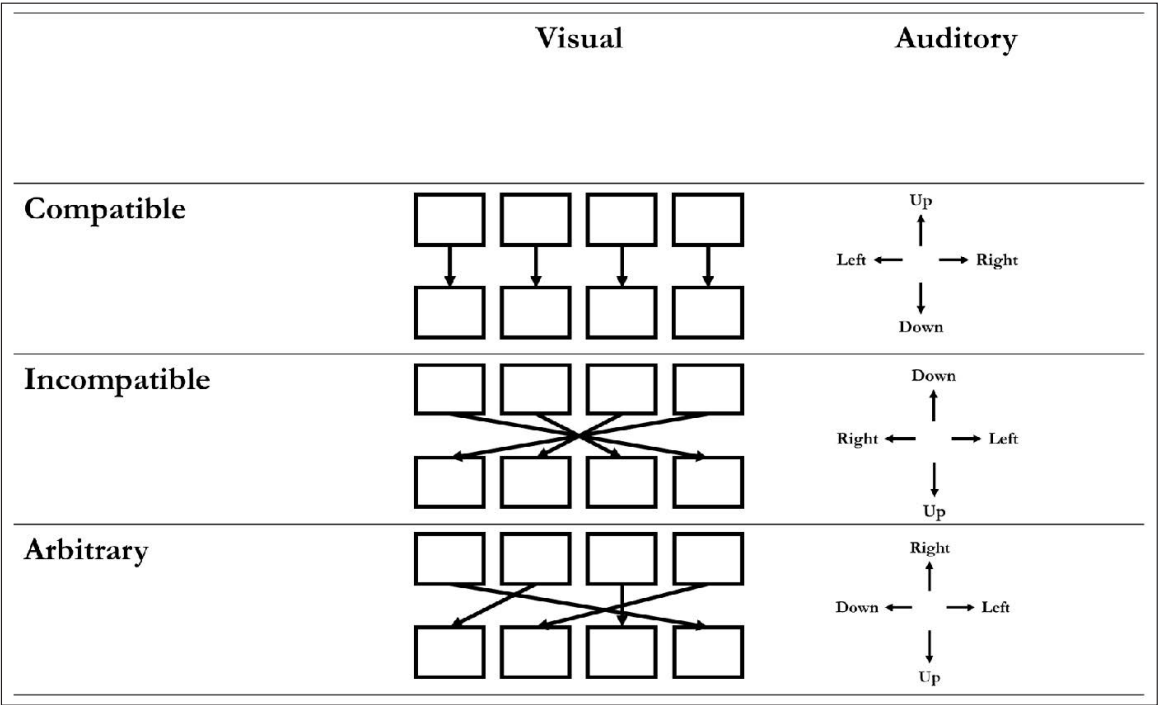


Figure 1: Stimulus-response sets for four-choice reaction time tasks in Wilhelm & Oberauer (2006).

Study 1
Method

Participants and testing procedure. Data were collected in 2005 from 155 students attending 11th and 12th grades at a school in Brandenburg, Germany in three 90-minute sessions. We dropped data from one participant not completing the RT measures, four participants with missing data in measures of WM and gf, and 15 participants performing below chance level in one or more RT tasks. The final sample included 135 participants (51 male, 84 female; age: mean = 17.5 years, SD = 0.7, range = 16–19).

Measures. An overview of measures in Study 1 can be found on the website of the Open Science Framework (OSF) (<https://osf.io/7y4wq/>). We used two-choice RT tasks with varying *SR-compatibility* (compatible vs. arbitrary) and *SR sets* (arrows vs. words vs. shapes) administered on a customized keyboard (picture available here: <https://osf.io/4ep6u/>). SR mappings are shown in **Figure 2**. An instruction emphasizing either speed or accuracy was applied to all six tasks, resulting in 12 conditions in total. Study 1 was part of a project that aimed to examine individual differences in conflict monitoring (see Keye, Wilhelm, Oberauer, & van Raavenzwaaij, 2009). The speed vs. accuracy instruction was relevant in this context. Choice RT tasks started with an instruction on the relevant SR mapping, followed by practice and test blocks containing trials of the same SR mapping. Stimuli remained on the screen until participants responded. We used two and four practice blocks for compatible and arbitrary conditions, respectively, with the exception of the first compatible and arbitrary block, which were preceded by four and eight practice blocks, respectively. During all practice blocks, participants received immediate feedback on the accuracy of their response, and a reminder cue of the mapping in use was displayed in the lower corner of the screen. For each condition of the RT tasks, we used 10 test blocks consisting of 9 trials each. After each practice and test block, feedback on mean accuracy and reaction time was provided. Data for the RT tasks including latency and accuracy for each trial are provided on the OSF website (<https://osf.io/pcz5q/>). The practice trials and the first trial in each block, RTs associated with erroneous responding, and RTs below 100 ms were discarded in further analyses. RTs were averaged for each of the 12 conditions. The accuracy for each condition was defined as proportion of correct (vs. incorrect) responses. **Table 1** provides an overview of descriptive statistics. We used five tests to measure *gf*, including four computerized tests (arrow series: Roberts & Stankov, 2001; 15 items from Set II of the Advanced Progressive Matrices: Raven, 1958; propositional reasoning: Wilhelm, 2005; number series: Wilhelm, 2005) and a composite based on six tasks (number series, ZN; figure analogy, AN; word analogy, WA; distinguishing between fact and opinion, TM; mathematical estimation, SC, and Charkov figure, CH) from


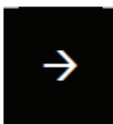












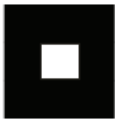





	Arrow		Word		Shape	
Compatible						
Stimulus						
Response			Left	Right		
Arbitrary						
Stimulus						
Response			Left	Right		

Figure 2: Stimulus-response sets for two-choice reaction time tasks in Study 1.

the Berlin Intelligence Structure test (BIS, Jäger, Süß, & Beauducel, 1997). Data for the gf tasks including scores on six BIS tasks as well as proportion-correct scores for arrow series, Raven's matrices, propositional reasoning, and number series respectively are provided on the OSF website (<https://osf.io/yrctdm/>). WM was measured with three tasks: rotation span (Shah & Miyake, 1996; Kane, Hambrick, Tuholski, Wilhelm, Pane, & Engle, 2004; Wilhelm & Oberauer, 2006), counting span (Engle, Kane, & Tuholski, 1999; Kane et al., 2004; Wilhelm & Oberauer, 2006), and memory updating (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Wilhelm & Oberauer, 2006). A description of the WM tasks (<https://osf.io/srtgn/>) as well as data including latency and accuracy for each item (<https://osf.io/yd4gk/>) are provided on the OSF website.

Data analysis. Experimental effects were tested with repeated measure ANOVAs. Correlations were examined in structural equation models. All variables were z-standardized before analysis. Models were estimated in R (R Core Team, 2017) with the lavaan package (Rosseel, 2012). Model fit was assessed by several criteria, including Comparative Fit Index (CFI; Bentler, 1990) higher than .95, Root-Mean-Square-of-Error-Approximation (RMSEA; Steiger, 1990) less than .08, and Standardized Root Mean Square Residual (SRMR) less than .08 (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996).

Results

Experimental effects. We conducted ANOVA with two within-subject variables: (1) instruction (speed vs. accuracy) and (2) SR mapping (compatible vs. arbitrary). Analyses were run separately by type of stimuli (arrow, shape, word). Results are summarized in **Table 2**. Central to our research questions, we observed effects of SR-compatibility on RT for two out of three tasks: The effect of SR-compatibility on RT was significant for arrow and shape tasks, but not for the word task. SR-compatibility had the expected effect on accuracy for all three tasks.

Measurement models for choice RT tasks. Because our main research questions concerned the effect of SR-compatibility on covariances of RT with gf and WM, we considered data from the word task not useful for answering these questions. Subsequent analyses were performed on RTs from arrow and shape tasks. We first tested a general factor model (Model A, see **Figure 3**) reflecting common variance among all choice RT tasks. Standardized model parameters and fit indicators are provided in **Figure 3**. Model A showed poor fit to data. Next, we established a latent factor reflecting shared variance due to instruction (speed vs. accuracy). In this model (Model B, see **Figure 3**), the nested factor "instruction" reflected shared variance due to accuracy instruction. The introduction of the nested instruction factor led to a significant increase in model fit ($\Delta\chi^2 = 104.58$, $\Delta df = 4$, $p < .01$), and to an improvement in fit indices. In a third step, we added a latent

Table 1: Descriptive Statistics for Latencies and Accuracies in the Choice Reaction Time Tasks in Study 1.

	Latencies						Accuracies			
	Mean	SD	Min	Max	Skewness	Kurtosis	Mean	SD	Min	Max
Arrow task										
Compatible, speed instruction	314	35	239	423	0.12	-0.02	.82	0.10	.50	1.00
Compatible, accuracy instruction	360	33	287	474	0.66	0.70	.96	0.04	.81	1.00
Arbitrary, speed instruction	312	44	208	441	0.19	0.24	.77	0.11	.50	1.00
Arbitrary, accuracy instruction	412	61	296	594	0.80	0.14	.96	0.04	.78	1.00
Shape task										
Compatible, speed instruction	240	20	186	312	0.34	0.70	.86	0.09	.62	1.00
Compatible, accuracy instruction	286	39	217	460	1.49	3.21	.99	0.02	.92	1.00
Arbitrary, speed instruction	326	43	228	435	0.05	-0.16	.80	0.12	.51	1.00
Arbitrary, accuracy instruction	385	45	300	607	1.41	4.14	.96	0.03	.85	1.00
Word task										
Compatible, speed instruction	333	42	225	475	0.04	1.57	.79	0.10	.52	.99
Compatible, accuracy instruction	413	47	329	580	1.10	1.59	.96	0.04	.81	1.00
Arbitrary, speed instruction	330	48	225	548	0.62	2.23	.76	0.11	.50	.96
Arbitrary, accuracy instruction	414	44	337	579	0.92	1.01	.96	0.04	.79	1.00

Note: N = 135, M = mean, SD = standard deviation, Min = Minimum, Max = Maximum. Latencies in milliseconds, accuracies in proportion of correct responses.

Table 2: Experimental Effects on Latencies and Accuracies in Study 1: Results from Repeated Measure ANOVAs.

	Latencies				Accuracies					
	df (hypothesis)	df (error)	F	p	partial eta-squared	df (hypothesis)	df (error)	F	p	partial eta-squared
Arrow task										
SRC	1	134	76.80	<.01	.36	1	134	40.84	<.01	.23
Instruction	1	134	448.64	<.01	.77	1	134	383.67	<.01	.74
SRC × Instruction	1	134	146.46	<.01	.52	1	134	29.53	<.01	.18
Shape task										
SRC	1	134	1394.70	<.01	.91	1	134	73.22	<.01	.35
Instruction	1	134	291.98	<.01	.69	1	134	333.91	<.01	.71
SRC × Instruction	1	134	8.94	<.01	.06	1	134	11.73	<.01	.08
Word task										
SRC	1	134	0.27	.60	.00	1	134	16.48	<.01	.11
Instruction	1	134	418.07	<.01	.76	1	134	515.17	<.01	.79
SRC × Instruction	1	134	0.88	.35	.01	1	134	14.32	<.01	.10

Note: N = 135. SRC = stimulus-response compatibility.

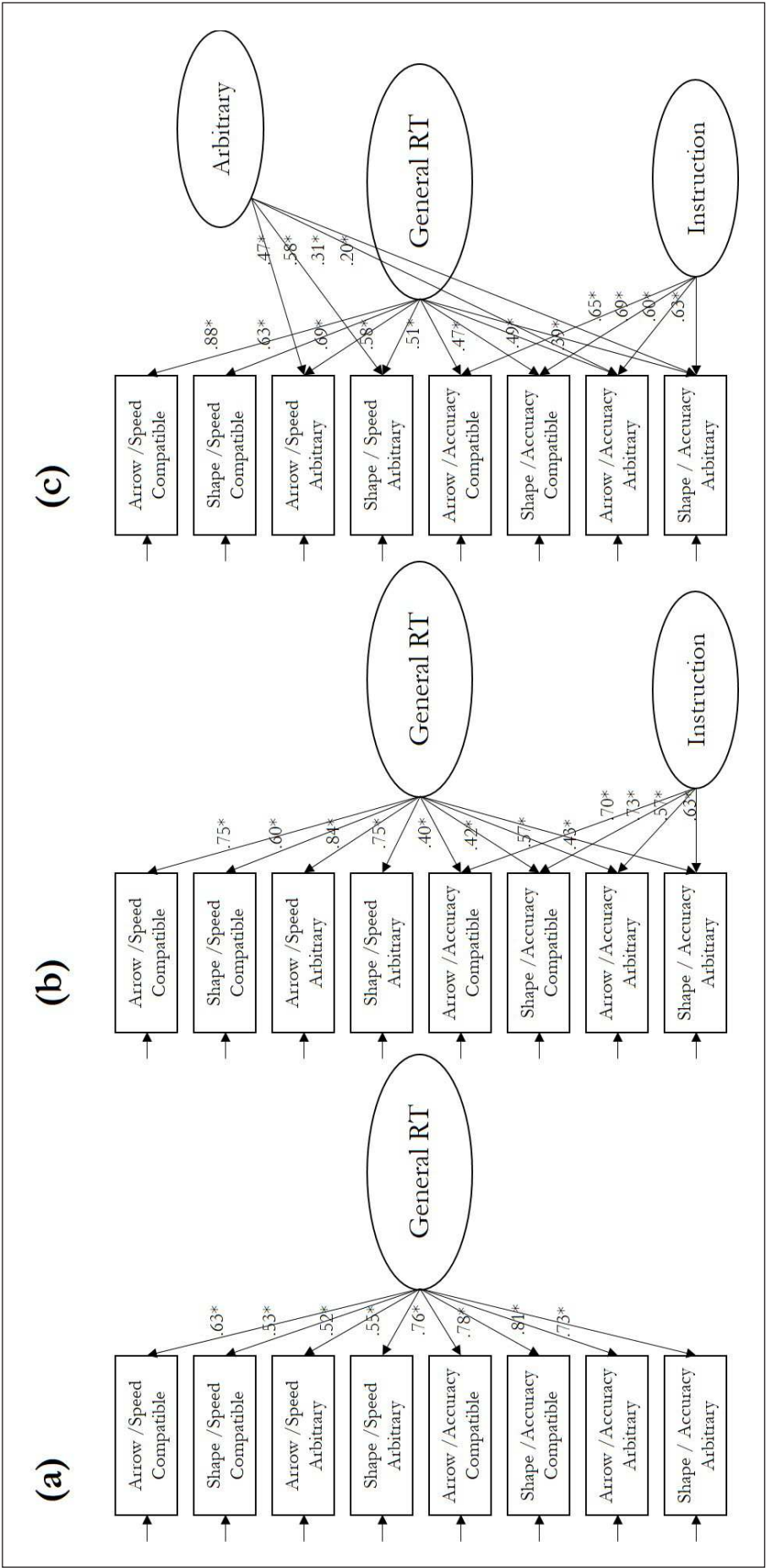


Figure 3: Measurement models for reaction time tasks in Study 1 and standardized parameter estimates. (a) Model A including a general reaction time factor. Model fit: $\chi^2[df] = 151.76[20]$, $p < .01$, CFI = .75, RMSEA = .22, SRMR = .12. (b) Model B with a nested factor reflecting the effects of the instruction (speed vs. accuracy). Model fit: $\chi^2[df] = 47.17[16]$, $p < .01$, CFI = .94, RMSEA = .12, SRMR = .06. (c) Model C with an additional nested factor reflecting binding costs in arbitrary conditions. Model fit: $\chi^2[df] = 19.92[12]$, $p = .07$, CFI = .99, RMSEA = .07, SRMR = .04. Models only included indicators of reaction time in the arrow and shape tasks because no consistent stimulus-response compatibility effects were found in the word task. Model B fit the data better than Model A and Model C fit the data better than Model B. $N = 135$, $p < .05$.

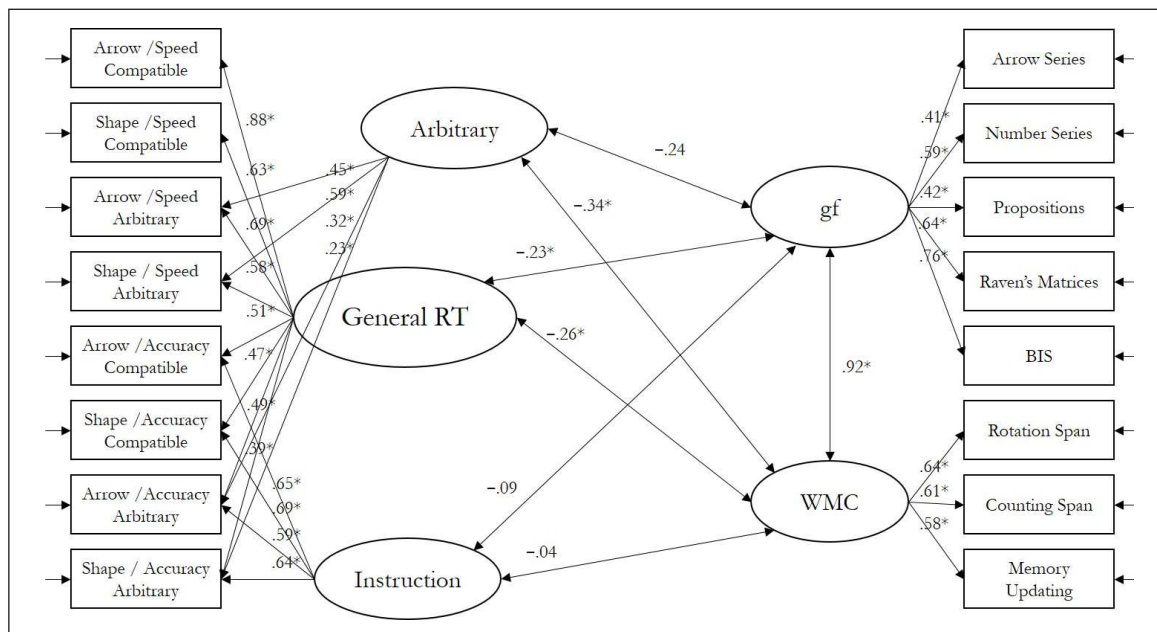


Figure 4: Structural model for examining associations of reaction time with fluid intelligence and working memory in Study 1. Model fit: $\chi^2 = 129.59$, $df = 89$, $p < .01$, CFI = 0.95, RMSEA = .06, SRMR = .06, $N = 135$, * $p < .05$.

“binding” factor reflecting shared variance due to arbitrary SR mappings (Model C, see **Figure 3**). This led to a further significant increase in model fit ($\Delta\chi^2 = 27.25$, $\Delta df = 4$, $p < .01$) and to an improvement in fit indices. All four loadings of the binding factor were significant (weakest loading: $p = .02$; all others $p < .01$).

Structural model. We extended Model C to test our hypotheses regarding correlations of RT with gf and WM. gf and WM were modeled as correlated factors. The structural model and standardized model parameters are shown in **Figure 4**. In line with the prediction of the binding hypothesis, the correlation of the binding factor with WM was significant ($\rho = -.34$, $p = .01$, CI = $-.59$ to $-.08$). The correlation between the binding factor and gf had the expected sign, but the 95% confidence interval included zero ($\rho = -.24$, $p = .06$, CI = $-.48$ to $.01$). General RT was correlated with gf ($\rho = -.23$, $p = .03$, CI = $-.44$ to $-.02$) and WM ($\rho = -.26$, $p = .02$, CI = $-.49$ to $-.04$). Correlations of the instruction factor with gf ($\rho = -.09$, $p = .39$, CI = $-.31$ to $.12$) and WM ($\rho = -.04$, $p = .71$, CI = $-.27$ to $.18$) were not significant. In sum, these findings are partially in line with those reported by Wilhelm & Oberauer (2006); however, correlations were smaller. Descriptive statistics for variables included in this model are provided in **Table 3**.

Follow-up analyses. In our structural model, the correlation between gf and WM was very high ($\rho = .92$, $p < .01$, CI = $.78$ to 1.05). Follow-up analyses indicated that a more parsimonious model with a general (g) factor did not fit the data worse ($\chi^2[df] = 28.32[20]$, $p = .10$, CFI = .96, RMSEA = .06, SRMR = .05) than the model with correlated gf and WM factors ($\chi^2[df] = 26.91[19]$, $p = .11$, CFI = .97, RMSEA = .06, SRMR = .05) as indicated by the χ^2 difference test ($\Delta\chi^2[\Delta df] = 1.42[1]$, $p = .23$). In a further step, we examined the correlations between general RT, binding, instruction, and the g factor. Faster general RT ($\rho = -.25$, $p = .02$, CI = $-.44$ to $-.05$) and lower binding costs ($\rho = -.28$, $p = .02$, CI = $-.51$ to $-.05$) were related to higher g, whereas the instruction factor did not correlate significantly with g ($\rho = -.08$, $p = .47$, CI = $-.28$ to $.13$).

Study 2

Method

Participants and testing procedure. Data were collected in 2007 and 2008 from 171 students attending 11th and 12th grades at 6 schools in Berlin and Brandenburg, Germany in two 90-minute sessions. We excluded data from 14 participants not completing both sessions and four participants performing below chance level at one or more RT tasks. The final sample included 153 participants (65 male, 88 female, age: mean = 17.2 years, SD = 0.9, range = 16–20).

Measures. An overview of measures in Study 2 can be found on the OSF website (<https://osf.io/cx52g/>). Two- and four-choice RT tasks were used with varying SR- compatibility (compatible vs. arbitrary) and SR

Table 3: Descriptive Statistics and Correlations of Variables Included in the Structural Model (Study 1).

	<i>M</i>	<i>SD</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) ARR/CO/SP	314	35	1															
(2) SHA/CO/SP	240	20	.53*	1														
(3) ARR/AR/SP	312	44	.62*	.50*	1													
(4) SHA/AR/SP	326	43	.51*	.36*	.67*	1												
(5) ARR/CO/AC	360	33	.48*	.36*	.20*	.28*	1											
(6) SHA/CO/AC	286	39	.41*	.38*	.27*	.30*	.68*	1										
(7) ARR/AR/AC	412	61	.44*	.36*	.42*	.47*	.62*	.64*	1									
(8) SHA/AR/AC	385	45	.35*	.31*	.29*	.36*	.59*	.63*	.62*	1								
(9) Arrow Series	61	14	-.09	-.14	-.11	-.12	.02	.05	-.07	-.06	1							
(10) Number Series	84	17	.01	-.20*	.02	-.15	-.09	-.18*	-.20*	-.25*	.16	1						
(11) Propositions	40	15	-.07	-.07	-.19*	-.14	-.01	-.03	-.07	-.05	.07	.21*	1					
(12) Raven's Matrices	57	18	-.11	-.19*	-.08	-.21*	-.01	.03	-.15	-.09	.32*	.31*	.37*	1				
(13) BIS	50	13	-.25*	-.20*	-.14	-.23*	-.19*	-.10	-.23*	-.22*	.30*	.47*	.33*	.49*	1			
(14) Rotation Span	75	15	-.18*	-.22*	-.19*	-.22*	.00	.01	-.20*	-.24*	.28*	.33*	.20*	.45*	.36*	1		
(15) Counting Span	85	12	-.07	-.15	-.19*	-.17*	-.01	-.04	-.10	-.11	.27*	.39*	.26*	.32*	.41*	.42*	1	
(16) Memory Updating	65	11	-.19*	-.19*	-.14	-.22*	-.13	-.16	-.17*	-.28*	.26*	.38*	.15	.27*	.45*	.39*	.31*	1

Note. *N* = 135, *M* = mean, *SD* = standard deviation. Variables were z-standardized before the analysis. ARR = arrow, SHA = shape, CO = compatible, AR = arbitrary, gf = fluid intelligence, WMC = working memory capacity. Latencies in milliseconds (Variables 1 to 8). gf and WM performance in percentage correct responses (variables 9 to 16). Correlations equal to or greater than |.17| are significantly different from 0 at *p* < .05.

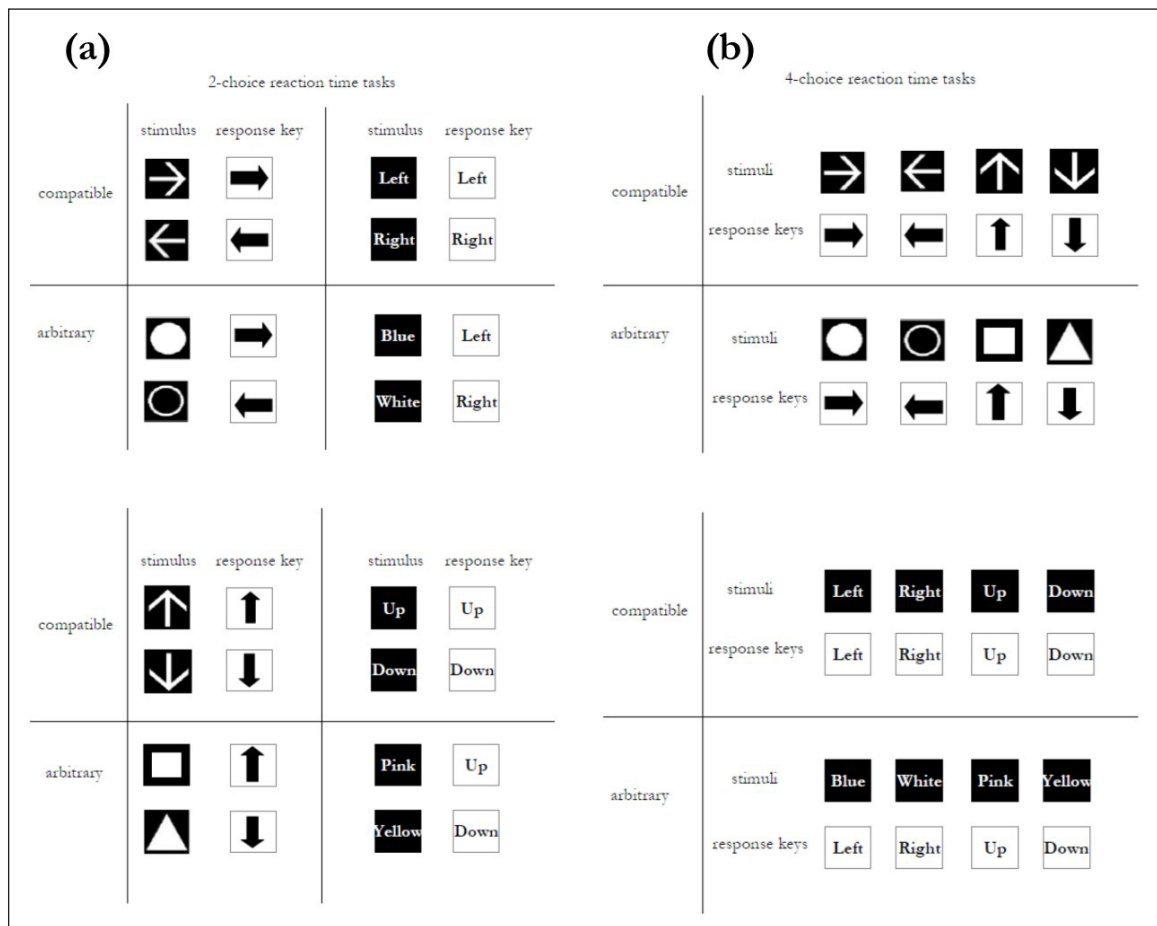


Figure 5: Stimulus-response sets for (a) two-choice and (b) four-choice reaction time tasks in Study 2.

sets (arrows vs. words) and administered on a customized keyboard (picture available here: <https://osf.io/tf4mc/>). Participants were asked to respond as fast and as accurately as possible. The shape task was not used in Study 2 because of time constraints. SR mappings are shown in **Figure 5**. The tasks started with an instruction on the relevant SR mapping followed by practice and test blocks containing trials of the same SR mapping. Stimuli remained on the screen until participants responded. Compatible and arbitrary conditions of two-choice tasks were practiced in two and four blocks consisting of 8 trials each, respectively. Compatible and arbitrary conditions of four-choice tasks were practiced in three and six blocks consisting of 16 trials each, respectively. For each condition of two-choice and four-choice tasks, we used 6 blocks consisting of 21 and 41 trials each, respectively. After each practice and test block, feedback on mean accuracy and RT was provided. Data for the RT tasks including latency and accuracy for each trial are provided on the OSF website (<https://osf.io/x3vkc/>). The practice trials and the first trial in each block were discarded in further analyses. Data treatment followed the same procedure as in Study 1. **Table 4** provides an overview of descriptive statistics for latencies and accuracies in choice RT tasks. We used four tests to measure *gf*, including three paper-pencil tests (solving equations, propositional reasoning, and matrices; Wilhelm, 2005) and a computerized version of 15 items from Set II of the Advanced Progressive Matrices (Raven, 1958). *gc* was measured with 43 items of the general knowledge test from the IST-2000-R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). Data for the *gf* and *gc* tasks including raw responses to each item are provided on the OSF website (<https://osf.io/rb7qu/>). Three composites were built for verbal, numeric, and figural content. WM was not measured because of time constraints.

Results

Experimental effects. We conducted ANOVA with two within-subject variables: (1) number of response alternatives (two-choice vs. four-choice) and (2) SR mapping (compatible vs. arbitrary). Analyses were run separately for both types of the word and arrow task (response keys: left/right and up/down and corresponding arrow keys, respectively). Results are summarized in **Table 5**. Central to our research question, SR-

Table 4: Descriptive Statistics for Latencies and Accuracies in the Choice Reaction Time Tasks in Study 2.

	Latencies						Accuracies			
	Mean	SD	Min	Max	Skewness	Kurtosis	Mean	SD	Min	Max
Arrow task										
2-choice (left-right) compatible	357	29	293	555	2.01	11.82	.93	0.05	.70	1.00
2-choice (left-right) arbitrary	399	44	282	563	0.72	1.29	.91	0.06	.63	1.00
2-choice (up-down) compatible	365	31	303	490	0.83	1.23	.92	0.06	.68	1.00
2-choice (up-down) arbitrary	391	40	316	555	0.90	1.57	.91	0.06	.57	.99
4-choice compatible	431	42	355	569	0.66	0.21	.94	0.05	.53	1.00
4-choice arbitrary	588	93	445	914	1.41	2.11	.90	0.06	.60	.99
Word task										
2-choice (left-right) compatible	402	33	334	499	0.56	0.11	.89	0.07	.58	.99
2-choice (left-right) arbitrary	400	39	312	572	0.95	1.90	.90	0.06	.65	.99
2-choice (up-down) compatible	422	41	303	539	0.36	0.16	.89	0.06	.60	1.00
2-choice (up-down) arbitrary	423	43	330	599	0.75	1.15	.89	0.06	.63	.99
4-choice compatible	565	66	452	884	1.29	3.10	.89	0.08	.41	.98
4-choice arbitrary	645	95	497	936	1.03	0.68	.88	0.09	.42	.98

Note: $N = 153$, M = mean, SD = standard deviation, Min = minimum, Max = maximum, Latencies in milliseconds, accuracies in proportion of correct responses.

compatibility had an effect on average latency. Effect sizes were stronger for arrow than word tasks. For both arrow and word tasks, SR-compatibility effects differed by the number of response alternatives, as indicated by the interaction effect. An examination of descriptive statistics (**Table 4**) suggested that for the arrow task, SR-compatibility had effects in both the two-choice and four-choice task, but that the effect was stronger for the four-choice task. For the word task, no effect of SR-compatibility on latencies was found for the two-choice tasks, while SR-compatibility had an effect on latencies in the four-choice task. Effects on accuracies were less consistent: SR-compatibility had at best minor effects on accuracy in the word task.

Measurement models for choice RT tasks. We only included RTs from the arrow task in the measurement model, because (a) SR-compatibility effects on RT were weaker in the word task, (b) SR-compatibility had no consistent effect on accuracies in the word task, and (c) this allowed keeping models more consistent across Studies 1 and 2. We first tested a general factor model reflecting common variance among all choice RT tasks with correlated residuals between compatible and arbitrary conditions of the four-choice task (Model A, see **Figure 6**). Standardized model parameters are provided in **Figure 6**. Model A showed acceptable fit by some of the fit criteria, except for RMSEA, which indicated poor fit. Next, we established a latent factor reflecting additional costs of binding (Model B, see **Figure 6**). The nested factor “binding” reflected shared variance due to arbitrary mappings. The introduction of the binding factor led to a significant increase in model fit ($\Delta\chi^2 = 39.66$, $\Delta df = 3$, $p < .01$) and to an improvement in fit indices. All loadings of the binding factor were significant ($p < .01$).

Structural model. We extended Model B to test our hypotheses regarding the correlation between RT and gf. The structural model and standardized model parameters are shown in **Figure 7**. The correlation between general RT and gf was moderately negative ($\rho = -.23$, $p = .02$, $CI = -.42$ to $-.03$). Contrary to the predictions of the binding hypothesis, the correlation between gf and binding was not significant ($\rho = -.08$, $p = .54$, $CI = -.31$ to $.17$). gc was unrelated to general RT ($\rho = -.12$, $p = .25$, $CI = -.31$ to $.08$) and binding ($\rho = -.02$, $p = .88$, $CI = -.26$ to $.22$). Descriptive statistics for variables included in this model are provided in **Table 6**.

Discussion

We examined effects of SR-compatibility on correlations of RT with gf and WM. Based on the binding hypothesis, we expected that gf and WM would be related to the ability to establish and uphold arbitrary bindings between independent elements. This ability is especially important with arbitrary SR mappings because there are no preexisting associations between stimuli and the corresponding responses. In Study 1, partially in line with the predictions of the binding hypothesis, WM was correlated with the latent binding factor ($\rho = -.34$, $p = .01$, $CI = -.59$ to $-.08$). The correlation between the binding factor and gf had the expected direction but was not significant ($\rho = -.24$, $p = .06$, $CI = -.48$ to $.01$). In Study 2, there was no evidence for a correlation of gf with the binding factor ($\rho = -.08$, $p = .54$, $CI = -.31$ to $.17$).

The present studies were conceptual replications of the study by Wilhelm and Oberauer (2006) who found higher correlations of WM and gf with a latent binding factor than with a general RT factor. Study 1 partially replicated these findings; however, the correlations with binding were substantially weaker ($\rho = -.24$ with gf, and $\rho = -.34$ with WM) than in the previous study ($\rho = -.55$ with gf, and $\rho = -.89$ with WM; Wilhelm & Oberauer, 2006). Study 2 did not replicate their findings. In sum, we found mixed evidence for the binding hypothesis and we could not replicate the strong correlation found by Wilhelm and Oberauer (2006). The study by Wilhelm and Oberauer (2006) was based on a smaller sample, and their correlations of binding with gf and WM were probably overestimated.

The inconsistency of the present results with Wilhelm and Oberauer (2006) is arguably due to a confound between response conflict and binding in the earlier study. One difference between the studies is that there was dimensional overlap between the SR sets both in compatible and arbitrary conditions in the study by Wilhelm and Oberauer (2006). In arbitrary conditions, the corresponding response key for each stimulus was present among the response alternatives, but not the correct response. This possibly resulted in response conflicts, which may have led to an increased need for inhibitory control (Kane & Engle, 2003). Therefore, individual differences in the ability to control response conflict (for instance through inhibition of erroneous response tendencies) offer an alternative explanation of the correlation of arbitrary SR mappings with WM and gf. Obviously, both explanations might also work in concert. In the current study, response conflicts were avoided by generating SR sets without dimensional overlap for the arbitrary-mapping conditions. One inevitable downside of the present task design is that the tasks with compatible vs. arbitrary mappings used different stimulus categories (e.g., arrows vs. colors). The compatibility effect is therefore potentially confounded with differences in how easily these stimulus categories can be perceptually identified and discriminated. Moreover, if the differences between stimuli vary across individuals, then individual differences

Table 5: Experimental Effects on Latencies and Accuracies: Results from Repeated Measure ANOVAs in Study 2.

	Latencies				Accuracies					
	df (hypothesis)	df (error)	F	p	Partial eta-squared	df (hypothesis)	df (error)	F	p	Partial eta-squared
Arrow task (left/ right)										
SRC	1	152	574	<.01	.79	1	152	84.48	<.01	.36
# response alternatives	1	152	1267	<.01	.89	1	152	0.35	.56	.00
SRC × # response alternatives	1	152	323	<.01	.68	1	152	16.18	<.01	.10
Arrow task (up/down)										
SRC	1	152	527	<.01	.78	1	152	70.62	<.01	.32
# response alternatives	1	152	1402	<.01	.90	1	152	2.59	.11	.02
SRC × # response alternatives	1	152	430	<.01	.74	1	152	32.69	<.01	.18
Word task (left/ right)										
SRC	1	152	148	<.01	.49	1	152	0.07	.79	.00
# response alternatives	1	152	2146	<.01	.93	1	152	4.71	.03	.03
SRC × # response alternatives	1	152	252	<.01	.62	1	152	16.84	<.01	.10
Word task (up/down)										
SRC	1	152	164	<.01	.52	1	152	5.47	.02	.03
# response alternatives	1	152	1845	<.01	.92	1	152	0.38	.54	.00
SRC × # response alternatives	1	152	221	<.01	.59	1	152	4.77	.03	.03

Note: N = 153, SRC = stimulus-response compatibility.

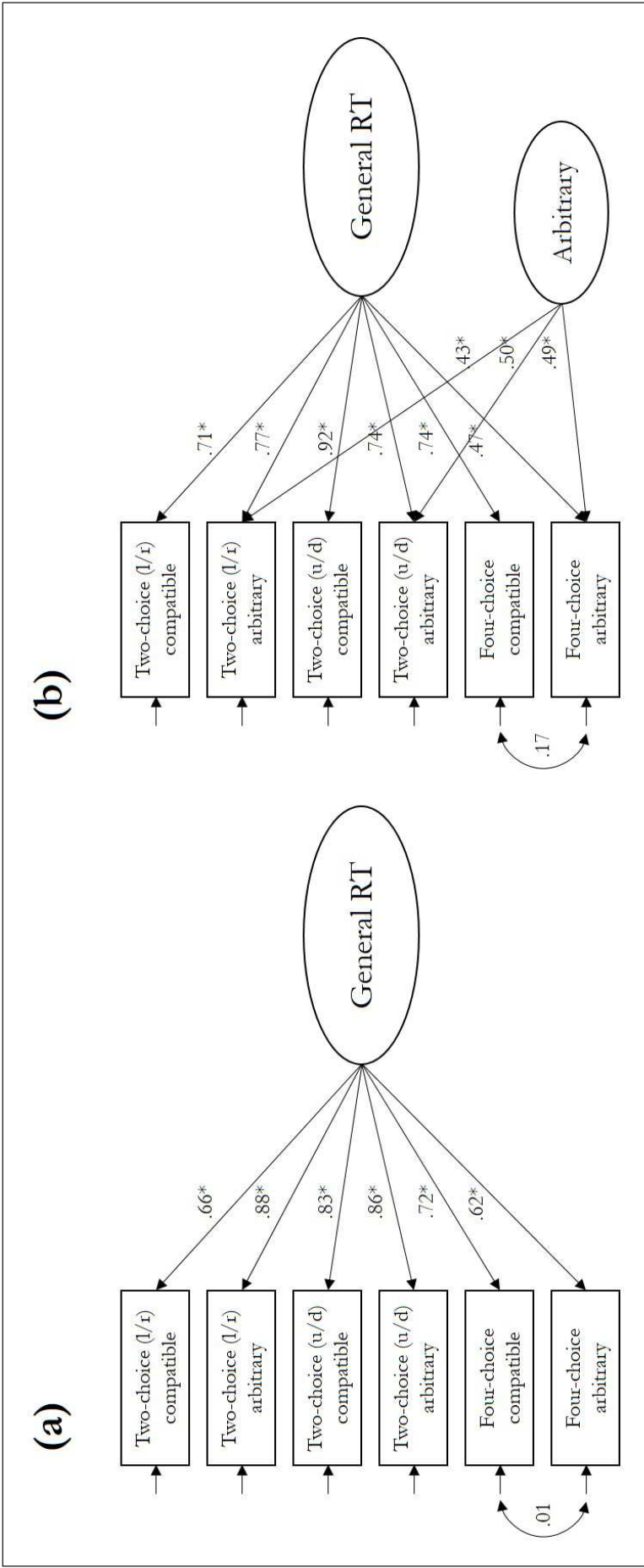


Figure 6: Measurement models for reaction time tasks in Study 2. (a) Model A including a general reaction time factor. Model fit: $\chi^2[df] = 43.14[8]$, $p < .01$, CFI = .93, RMSEA = .17, SRMR = .05. (b) Model B with a nested factor reflecting binding costs in arbitrary conditions. Model fit: $\chi^2[df] = 3.49[5]$, $p = .63$, CFI = 1.00, RMSEA = .00, SRMR = .01. Models only included indicators of reaction time in the arrow task. Model B fit the data better than Model A. $N = 153$, * $p < .05$.

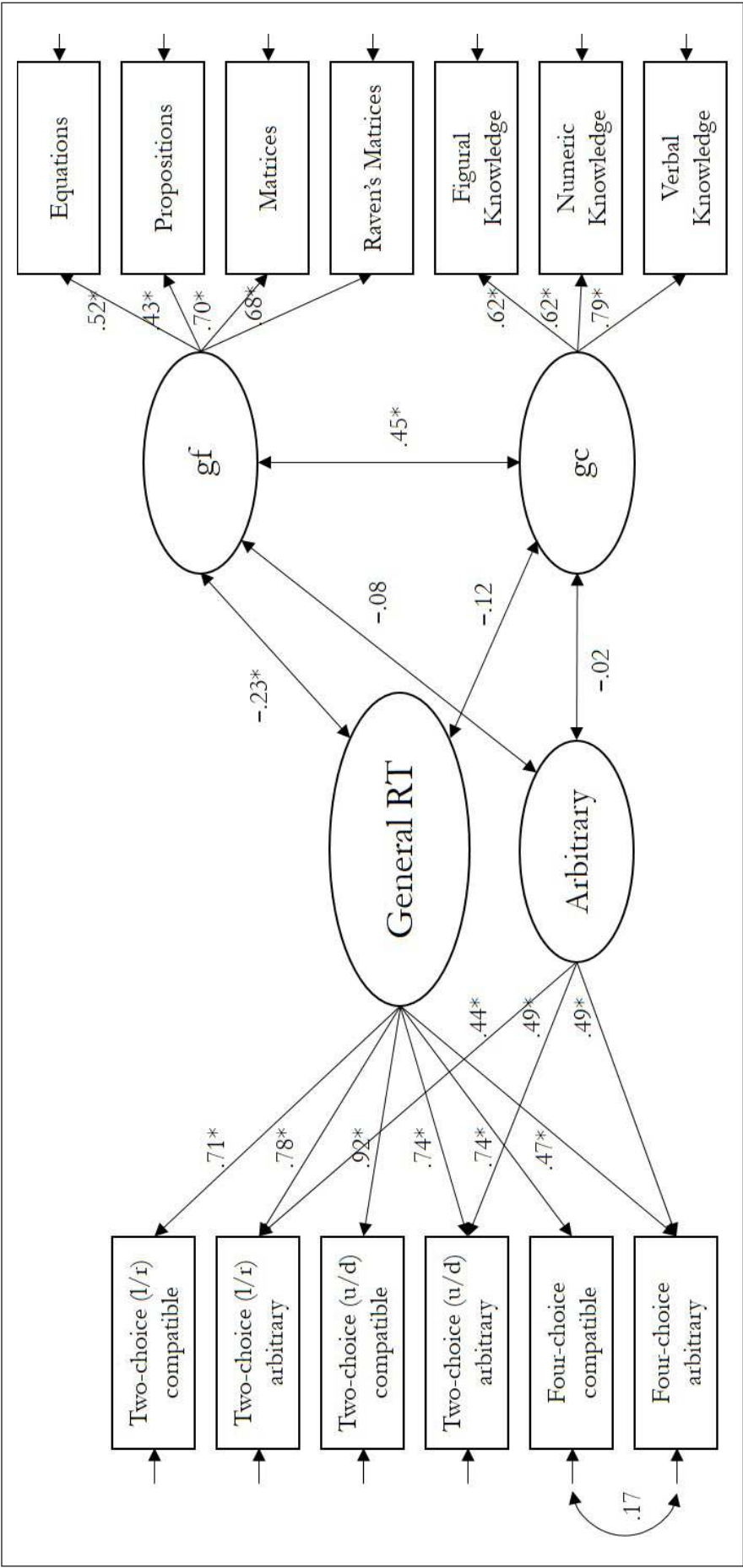


Figure 7: Structural model for examining associations between reaction time and intelligence in Study 2. Model fit: $\chi^2 = 72.11$, $df = 56$, $p = .07$, CFI = 0.98, RMSEA = .04, SRMR = .06, $N = 153$, $p < .05$.

Table 6: Descriptive Statistics and Intercorrelations of Variables Included in the Structural Model in Study 2.

	<i>M</i>	<i>SD</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) LR/CO	357	29	1												
(2) LR/AR	399	44	.55*	1											
(3) UD/CO	365	31	.66*	.71*	1										
(4) UD/AR	391	40	.51*	.79*	.67*	1									
(5) 4C/CO	431	42	.51*	.57*	.68*	.60*	1								
(6) 4C/AR	588	93	.32*	.57*	.42*	.59*	.44*	1							
(7) gf: Equations	56	22	-.18*	-.25*	-.18*	-.09	-.11	-.16*	1						
(8) gf: Propositions	48	22	-.18*	-.30*	-.20*	-.18*	-.22*	-.09	.17*	1					
(9) gf: Matrices	54	21	-.12	-.14	-.16	-.08	-.12	-.16*	.37*	.31*	1				
(10) gf: Raven's Matrices	57	20	-.04	-.14	-.02	.00	-.04	-.01	.33*	.28*	.49*	1			
(11) gc: Figural	62	13	-.17*	-.13	-.08	-.06	-.12	-.03	.25*	.14	.17*	.27*	1		
(12) gc: Numeric	56	19	.06	-.07	.07	-.01	.05	.00	.21*	.07	.22*	.23*	.34*	1	
(13) gc: Verbal	65	13	-.11	-.12	-.10	-.03	-.12	-.05	.18*	.16	.15	.25*	.49*	.50*	1

Note: *N* = 153. Variables were z-standardized before the analysis. *M* = mean, *SD* = standard deviation, LR = left/right, UD = up/down, CO = compatible, AR = arbitrary, gf = fluid intelligence, gc = crystallized intelligence. Latencies in milliseconds. gf and gc performance in percentage correct responses. * $p < .05$.

in compatibility effects could in part reflect individual differences in perceptual identification and discrimination. This contamination of the variance reflected in the “binding” factors could be responsible for their weaker associations with intelligence and WM capacity. Against this assumption, Meier & Kane (2015) did not find any evidence that RT costs due to SR incompatibility were related to individual differences in WM. However, their study involved compatible vs. incompatible, but no arbitrary SR mappings.

Another important difference between the present studies and the study by Wilhelm & Oberauer (2006) is that the previous study involved a more heterogeneous sample. The participants in the present studies were 11th and 12th grade students, whereas the participants in Wilhelm & Oberauer ranged between 18 and 36 years in age. They were also more heterogeneous in terms of educational background: 82% of participants had completed the German academic high school track, whereas all participants in the present studies were students attending the academic track. This may have resulted in higher correlations between all cognitive variables in Wilhelm & Oberauer. Specifically, we found a correlation of $\rho = .45$ between gf and gc in Study 2 whereas Wilhelm and Oberauer reported a correlation of $\rho = .75$. Furthermore, the correlations between Arrow Series, Number Series, Propositions, and Raven's matrices in Study 1 ranged from $r = .07$ to $.37$. The same correlations ranged from $r = .32$ to $.50$ in Wilhelm and Oberauer (2006). The correlations between rotation span, counting span, and memory updating in Study 1 ranged from $r = .31$ to $.42$; the same correlations ranged from $r = .47$ to $.57$ in Wilhelm & Oberauer (2006). In contrast, the correlation between gf and WM was somewhat higher in Study 2 ($\rho = .92$) than in the study by Wilhelm & Oberauer (2006) ($\rho = .81$).

In studies 1 and 2, we observed that SR-compatibility depended on the SR set: With words as stimuli, no or weaker effects of SR-compatibility were found. More research is needed to understand why effects of SR-compatibility were non-existent or weaker in the word task.

Summary and conclusion

The studies reported here were conceptual replications of Wilhelm & Oberauer (2006). In sum, the findings provided mixed evidence for the binding hypothesis. In Study 1, findings were partially in line with the binding hypothesis, but correlations were weaker than those reported by Wilhelm and Oberauer (2006). The findings of Study 2 did not provide evidence for the binding hypothesis. More research is needed to be able to separate different processes (e.g., response conflict, binding, perceptual identification and discrimination) involved in choice RT and to understand associations with individual differences in intelligence and WM.

Data Accessibility Statement

Data and code available at: <https://osf.io/hdyfq/>.

Ethics and Consent

At the time the data were collected (2005–2008), neither the funding organization (German Research Foundation DFG) nor the university at which the research was conducted (Humboldt University, Berlin, Germany) required ethical approval for this kind of purely behavioral research.

Acknowledgements

We thank Benjamin Goecke for assistance with data preparation.

The authors acknowledge the support of the German Research Foundation (grant number Wi2667/4) and the Open Access Publishing Fund for Social Sciences and Humanities at the University of Zurich.

Competing Interests

The authors have no competing interests to declare.

References

- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A.** (2001). *Intelligenz-Struktur-Test 2000 R [Intelligence Structure Test 2000 R]*. Göttingen: Hogrefe.
- Bentler, P. M.** (1990). Comparative fit indexes in structural models. *Psychological bulletin*, 107, 238–246. DOI: <https://doi.org/10.1037/0033-2909.107.2.238>
- Danthiir, V., Roberts, R. D., Schulze, R., & Wilhelm, O.** (2005). Mental speed: On frameworks, paradigms, and a platform for the future. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence*, (pp. 27–46). London: Sage. DOI: <https://doi.org/10.4135/9781452233529.n3>

- Engle, R. W., Kane, M. J., & Tuholski, S. W.** (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 102–134). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139174909.007>
- Fitts, P. M., & Deininger, R. L.** (1954). S-R compatibility: correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology*, 48, 483–492. DOI: <https://doi.org/10.1037/h0054967>
- Fitts, P. M., & Seeger, C. M.** (1953). S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46, 199–210. DOI: <https://doi.org/10.1037/h0062827>
- Hu, L. T., & Bentler, P. M.** (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. DOI: <https://doi.org/10.1080/10705519909540118>
- Jäger, A. O., Süß, H.-M., & Beauducel, A.** (1997). *Berliner Intelligenzstruktur - Test. Form 4. [Berlin Intelligence Structure Test: Form 4]*. Göttingen: Hogrefe.
- Jensen, A. R.** (1993). Spearman's g: Links between psychometrics and biology. *Annals of the New York Academy of Sciences*, 702, 103–129. DOI: <https://doi.org/10.1111/j.1749-6632.1993.tb17244.x>
- Kane, M. J., & Engle, R. W.** (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology General*, 132, 47–70. DOI: <https://doi.org/10.1037/0096-3445.132.1.47>
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. E.** (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189–217. DOI: <https://doi.org/10.1037/0096-3445.133.2.189>
- Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D.** (2009). Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Erikson flanker task. *Psychological Research*, 73, 762–776. DOI: <https://doi.org/10.1007/s00426-008-0188-9>
- Kornblum, S., Hasbroucq, T., & Osman, A.** (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility – a model and taxonomy. *Psychological Review*, 97, 253–270. DOI: <https://doi.org/10.1037/0033-295X.97.2.253>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M.** (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149. DOI: <https://doi.org/10.1037/1082-989X.1.2.130>
- Meier, M. E., & Kane, M. J.** (2015). Carving executive control at its joints: Working memory capacity predicts stimulus-stimulus, but not stimulus-response, conflict. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 1849–1872. DOI: <https://doi.org/10.1037/xlm0000147>
- Meiran, N., Pereg, M., Givon, E., Danieli, G., & Shahar, N.** (2016). The role of working memory in rapid instructed task learning and intention-based reflexivity: An individual differences examination. *Neuropsychologia*, 90, 180–189. DOI: <https://doi.org/10.1016/j.neuropsychologia.2016.06.037>
- Meiran, N., & Shahar, N.** (2018). Working memory involvement in reaction time and its contribution to fluid intelligence: An examination of individual differences in reaction-time distributions. *Intelligence*, 69, 176–185. DOI: <https://doi.org/10.1016/j.intell.2018.06.004>
- Oberauer, K.** (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, 134, 368–287. DOI: <https://doi.org/10.1037/0096-3445.134.3.368>
- Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W.** (2000). Working memory capacity: Facets of a cognitive ability construct. *Personality and Individual Differences*, 29, 1017–1045. DOI: [https://doi.org/10.1016/S0191-8869\(99\)00251-2](https://doi.org/10.1016/S0191-8869(99)00251-2)
- Raven, J. C.** (1958). *Advanced progressive matrices*. (2nd ed.). London: Lewis.
- R Core Team.** (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Roberts, R. D., & Stankov, L.** (2001). *Omnibus Screening Protocol*. Sydney: E-ntelligent Testing Products.
- Rosseel, Y.** (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1–36. DOI: <https://doi.org/10.18637/jss.v048.i02>
- Shah, P., & Miyake, A.** (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27. DOI: <https://doi.org/10.1037/0096-3445.125.1.4>

- Steiger, J. H.** (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180. DOI: https://doi.org/10.1207/s15327906mbr2502_4
- Unsworth, N., Redick, T. S., Spillers, G. J., & Brewer, G. A.** (2012). Variation in working memory capacity and cognitive control: Goal maintenance and micro-adjustments of control. *Quarterly Journal of Experimental Psychology*, 65, 326–355. DOI: <https://doi.org/10.1080/17470218.2011.597865>
- Wilhelm, O.** (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Understanding and Measuring Intelligence* (pp. 373–392). London: Sage. DOI: <https://doi.org/10.4135/9781452233529.n21>
- Wilhelm, O., & Oberauer, K.** (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, 18, 18–50. DOI: <https://doi.org/10.1080/09541440500215921>

How to cite this article: Hülür, G., Keye-Ehing, D., Oberauer, K. and Wilhelm, O. 2019 The Effect of Stimulus-Response Compatibility on the Association of Fluid Intelligence and Working Memory with Choice Reaction Times. *Journal of Cognition*, 7(1): 14, pp. 1–19. DOI: <https://doi.org/10.5334/joc.66>

Submitted: 30 November 2018

Accepted: 06 May 2019

Published: 11 June 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[*Journal of Cognition* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 